An Interrater Reliability Study of Rorschach Performance Assessment System (R-PAS) Raw and Complexity-Adjusted Scores

Pignolo, Claudia; Giromini, Luciano; Ando', Agata; Ghirardello, Davide; Di Girolamo, Marzia; Ales, Francesca; Zennaro, Alessandro. An Interrater Reliability Study of Rorschach Performance Assessment System (R–PAS) Raw and Complexity-Adjusted Scores. JOURNAL OF PERSONALITY ASSESSMENT. None pp: 1-7.
DOI: 10.1080/00223891.2017.1296844

R-PAS INTER-RATER RELIABILITY

Abstract

Recently, the Rorschach Performance Assessment System (R-PAS; Meyer, Viglione, Mihura, Erard, & Erdberg, 2011) was introduced to overcome some possible limitations of the Comprehensive System (CS; Exner, 2003) while continuing its efforts to link Rorschach inferences to their evidence base. An important, technical modification to the scoring system is that R-PAS interpretations are based on both standard scores and complexity-adjusted scores. Two previous U.S. studies reported good to excellent inter-rater reliability (IRR) for the great majority of R-PAS variables; however, IRR of complexity-adjusted scores has never been investigated. Furthermore, no studies have yet investigated R-PAS IRR in Europe. To extend this literature, we examined R-PAS IRR of Page 1 and Page 2 raw and complexity-adjusted scores with 112 Italian Rorschach protocols. We collected a large sample of both clinical and nonclinical Rorschach protocols, each of which was coded separately by two independent raters. Results demonstrated a mean intraclass correlation of .78 (*SD* = .14) for raw scores and of .74 (*SD* = .14) for complexity-adjusted scores. Overall, for both raw and complexity-adjusted values, most of the variables were characterized by good to excellent IRR.

Keywords: R-PAS; Rorschach; Inter-rater reliability; Coding; Complexity.

An Inter-rater Reliability Study of Rorschach Performance Assessment System (R-PAS) Raw and Complexity-Adjusted Scores

In Rorschach-based, psychological assessment literature, inter-rater reliability (IRR) refers to the extent to which different raters would code a given Rorschach protocol consistently. Since 1990s, doubts about the reliability of the Rorschach Comprehensive System (CS; Exner, 2003) have been raised (Wood, Nezworski, & Stejskal, 1996). However, over the years numerous studies have reported good IRR for the CS, suggesting that well trained raters could code reliably (Acklin, McDowell, Verschell, & Chan, 2000; Exner, 1993; McDowell & Acklin, 1996; Meyer, 2004; Meyer & Archer, 2001; Viglione, 1999; Viglione & Meyer, 2008; Viglione & Taylor, 2003). In fact, the most recent meta-analysis on this matter (Meyer, 2004) indicates that IRR for many Rorschach variables produce intraclass correlation (ICC) values that may be characterized as good or excellent. According to Meyer (2004), "this level of agreement compares favorably with the reliability seen for a wide range of other determinations made in psychology and medicine" (p. 325).

Recently, the Rorschach Performance Assessment System (R-PAS; Meyer, Viglione, Mihura, Erard, & Erdberg, 2011) was introduced to overcome some of the psychometric limitations of the CS, while continuing its efforts to link Rorschach inferences to their evidence base. Compared to CS, R-PAS has introduced some important, technical modifications (Meyer, 2011; Meyer & Eblin, 2012). Among them, a new administration procedure has been implemented (see Reese, Viglione, & Giromini, 2014; Viglione et al., 2015), and new international norms have been developed (see Giromini, Viglione, & McCullaugh, 2015; Viglione & Giromini, 2016). Although most of the CS codes were incorporated into R-PAS, so

far only two studies have investigated IRR of R-PAS.

In the first of these studies, Viglione, Blume-Marcovici, Miller, Giromini, and Meyer (2012) reported data on two graduate students who independently coded 50 Rorschach protocols administered to adults and children. They found that the mean ICC was .88 (*SD* = .11), and the median was .92. Based on Cicchetti (1994) and Shrout and Fliess' (1979) interpretative benchmarks, none of the 60 R-PAS variables showed poor reliability (ICC < .40), and 90% of the ICCs indicated excellent reliability (ICC ≥ .75). Only two variables showed fair reliability, Vista (V; ICC = .44) and Vagueness Percent (Vg%; ICC = .54). The four variables that yielded good reliabilities were: Form Quality Unusual Percent (FQu%; ICC = .64), Form Dimension (FD; ICC = .66), Inanimate Movement (m; ICC = .69), and Color Dominance Proportion [(CF+F)/SumC; ICC = .72].

In the second study, Kivisalu, Lewey, Shaffer, and Canfield (2016) evaluated IRR for 50 nonclinical R-PAS protocols at response-level of analysis. Each protocol was coded twice, first by the original examiner and then by a blind coder. Overall, percent agreement was excellent, ranging from 82.7% to 100%, and the mean ICC of the 62 codes was .78 (range: .30 – 1.00). However, three codes showed poor ICC values, which were Vista (V; ICC = .32), Deviant Responses Level 1 (DR1; ICC = .30), and Peculiar Logic (PEC; ICC = .39). Taken together, the findings of both these studies suggest that the majority of R-PAS variables achieve good to excellent IRR, though some variables (e.g., Vista) tend to produce less optimal results.

Inter-rater Reliability of Raw and Complexity Adjusted Scores

In R-PAS, when interpreting a Rorschach protocol one must first take into consideration its overall level of "complexity," conceptually defined by Viglione (1999) as "the amount of productivity, precision, differentiation, and integration involved in the aggregate of all the

responses" (p. 259). More technically, complexity refers to the first factor of the Rorschach (Meyer, Viglione & Giromini, 2014), and so it reflects the shared variance in common with all test scores. Because it is correlated with many Rorschach variables and this extraneous variance may reduce interpretative validity, in addition to raw standard scores, R-PAS also offers complexity-adjusted, standard scores to establish what the score of an examinee would be if his or her level of complexity was at the median. These newly introduced scores take into account the contribution of Complexity to each score and indicate how much an examinee's observed score diverges from the expected value based on his or her Complexity level. For example, a person with an observed Human Movement score (M) of 4 is at an average level when disregarding Complexity; however, with a low level of Complexity a score of 4 could be considered higher than expected and, conversely, with a high level of Complexity, the same score could be considered lower than expected (see Meyer et al., 2011). Put simply, complexity-adjusted scores remove from a given score the impact that Complexity had on generating that raw value. Despite the innovative nature of these scores, the IRR of complexity-adjusted scores has never been investigated.

Viglione et al. (2012) reported an ICC value for Complexity of .99. Because of this nearly perfect IRR value, variables that are highly correlated with Complexity will necessarily tend to show high IRR values as well. Conversely, variables that are completely unrelated to Complexity may or may not produce high IRR values. As such, because complexity-adjusted scores remove from Rorschach variables the effects of Complexity, the IRR of variables that are highly correlated with Complexity will likely tend to be higher when considering raw rather than complexity-adjusted scores. Indeed, computing ICCs for complexity-adjusted scores produced by two independent coders is conceptually similar to computing partial correlations for raw

scores produced by two independent coders, after controlling for Complexity. Because Complexity is coded with an almost perfect reliability (i.e., ICC ≈ 1), removing its effects from IRR analyses reduces the possibility that the raw scores of two independent coders look similar to each other just because they share a common variance with a third variable (i.e., Complexity), thus generating a sort of spurious relationship. Said differently, IRR analyses of complexity-adjusted scores eliminate the possibility that two raw scores show high ICCs just because they both correlate with a third variable that is coded with an extremely high reliability. Accordingly, when testing the IRR of R-PAS variables, one may anticipate that complexity-adjusted scores would produce either similar or lower ICCs than raw scores.

The Current Study

Given that R-PAS is gaining popularity among accredited U.S. doctoral training programs (Mihura, Roy, & Graceffo, 2016), and that to date only two U.S. studies have reported on its IRR – none of which presented data on complexity-adjusted scores – the current investigation attempted to examine R-PAS IRR of both raw and complexity-adjusted scores, with 112 Italian Rorschach protocols. Our study had two main objectives: (1) investigating the generalizability of U.S. IRR findings to data collected in Italy, and (2) evaluating whether complexity-adjusted scores would show lower level of reliability compared to raw scores that are computed disregarding Complexity.

Method

To evaluate IRR of R-PAS raw and complexity-adjusted scores, we collected a large sample of both clinical and nonclinical Rorschach protocols, each of which was coded by two independent raters. In line with Shrout and Fliess' (1979) guidelines, IRR of Page 1 and Page 2 R-PAS variables was calculated using one-way random effects model, intraclass correlation

coefficients (ICCs). Most of the studies on the IRR of Rorschach scores (e.g., Acklin et al., 2000; Viglione et al., 2012) used the two-way random effect model, which assumes that the same pair of raters have rated each protocol. In our study, however, the pair of raters was not the same for all protocols so that the one-way random effects model was preferable for our study (see Meyer et al., 2002).

Participants

A total of 112 Rorschach records were selected from archival clinical files and ongoing Rorschach studies available to the authors. Both clinical and nonclinical data were investigated. All protocols were collected using the R-Optimized administration (Meyer et al., 2011). The nonclinical subsample was composed of 44 (39% of the sample) college students, most of whom were women (80%), and the mean age was 21 years (*SD* = 1.6). The total number of responses of the nonclinical subsample was 1,158 with an average number of 26.3 (*SD* = 3.4) responses per protocol.

Among the clinical subsample, 29 (26% of the sample) were children with a diagnosis of Attention Deficit and Hyperactivity Disorders (ADHD), referred to public mental health services for psychological evaluation. Most of the children were females (79%), and the mean age was 11.8 years (*SD* = 2.7). The average number of responses was 23.3 (*SD* = 4.4) with a total of 675 responses. Among the 39 adults of the clinical subsample, 16 (14% of the sample) were women with a diagnosis of Rheumatoid Arthritis (RA), and 23 (21% of the sample) were women with a diagnosis of Fibromyalgia Syndrome (FMS). The Rorschach was administered to both groups for research purposes. For the RA group, the mean age was 54.8 years (*SD* = 11.1) and the total number of responses was 406, with an average of 25.4 (*SD* = 4.5) responses per protocol. The FMS group was composed of women with a mean age of 50.1 (*SD* = 9.5). The total number of

responses was 613, with an average of 26.7 (*SD* = 4.2) responses per protocol. The total number of responses coded by raters was 2,852, with an average of 25.5 (*SD* = 4.2) responses per protocol.

Rorschach coders

Five raters contributed to this study. Each of them contributed by independently coding all responses of a selected number of protocols, so that each protocol was eventually coded twice. That is, each protocol was coded by two different, independent raters. Four of the raters were graduate students, whereas the fifth rater had completed a doctoral program; all had been trained in both CS and R-PAS. Three raters had been trained by the same mentor (the second author), who had achieved administration and coding proficiency in 2013 and 2014 respectively. The other two raters had attended online R-PAS workshops and had been in training, for a brief amount of time, with Dr. Donald Viglione, one of the developers of R-PAS. At the time the data were being collected, however, none of the raters had yet achieved administration or coding proficiency by R-PAS.

Statistical Analysis

Before investigating IRR of R-PAS variables, we evaluated the normality of scores' distributions. We applied square root transformations to variables that departed substantially from normality (i.e., skewness > 2 and kurtosis > 7; West, Finch, and Curran, 1995). Moreover, because R-PAS proportion scores cannot be computed when the denominator is equal to zero (and individuals' scores cannot be used for computing IRR), we computed percentage scores as the difference between the numerator and the second code composing the denominator. For instance, the Human Movement Proportion (M/MC, which in R-PAS is obtained by dividing M by the sum of M and WSumC) was computed as Human Movement minus Weighted Sum of

Color (M – WSumC). Recently, the R-PAS authors have suggested this alternative procedure, in that using difference scores instead of proportions allows researchers to use all protocols in their dataset (www.r-pas.org).

Because the base rate may affect IRR (for details, see Viglione et al., 2012), we computed base rates for each variable. Several studies revealed that, generally, low base rate variables are characterized by lower reliability (e.g., Vista); given that low base rate variables occur less than once per protocol, raters may not have many occasions of practicing in coding those variables and so they may not code them reliably. To demonstrate the frequency with which each variable appeared in the dataset, we followed procedures discussed in Viglione et al. (2012), i.e., a mean frequency value lower than 1 was considered to be indicative of a "rare" base rate, mean frequency values between 1 and 2 were considered "infrequent", and mean frequency values greater than 2 were defined as "common". Finally, in evaluating ICCs, we followed the guidelines suggested by Cicchetti (1994) and Shrout and Fliess (1979): ICC values lower than .40 indicate poor reliability, between .40 and .59 fair reliability, between .60 and .74 good reliability, and values at or above .75 suggest excellent reliability. As noted above, both raw and complexity-adjusted scores were examined.

Results

The IRR results of Page 1 and Page 2 R-PAS variables are shown in Tables 1, 2, and 3. In Table 1, ICCs are reported for both raw and complexity-adjusted scores. As expected, Complexity showed an excellent IRR (ICC = .94), and ICCs for complexity-adjusted scores were slightly lower than ICCs for raw scores. As reported in Table 2, the mean ICC is .78 (*SD* = .14) for raw scores and .74 (*SD* = .14) for complexity-adjusted scores. The lowest levels of reliability (i.e., ICC < .50) for complexity-adjusted scores were observed in three Page 2 variables: Pure

Color (C; ICC = .41), Passive Human Movement minus Active Human Movement (Mp - Ma; ICC = .43), and Passive Movement minus Active Movement (p – a; ICC = .47). For raw scores, C (ICC = .41) had the lowest ICC. Overall, for both raw and complexity-adjusted values, 82% of the variables was characterized by good to excellent reliability (Cicchetti, 1994; Shrout & Fleiss, 1979).

Comparing the reliability results for the two different types of scores, i.e., raw and complexity-adjusted scores, we found small differences (mean ICC difference = .03, *SD* = .05). Differences greater than .10 (which is approximatively 1.5 *SD*'s above the mean) were observed for four variables only: Human Movement and Weighted Sum of Color (MC; difference = .27), Synthesis (Sy; difference = .19), Blend (difference = .15), and Human Movement (M; difference = .13). As expected, when looking at complexity-adjusted rather than raw scores, the ICCs of these variables tended to be smaller. However, in all these cases, the ICCs of the complexity-adjusted scores were still in the good to excellent interpretative range.

Lastly, we also evaluated the relationship between base rates and ICCs. Looking at base rates, among the 12 variables with low base rate, only three yielded fair ICCs: Human Movement responses with FQ Minus (M-), Vista, and C. Overall, the mean ICC for low base rate variables (*M* = .74 and *M* = .72 for raw and complexity-adjusted scores respectively) was higher than those for infrequent variables (*M* = .69 and *M* = .67 for raw and complexity-adjusted scores respectively).

Additional analyses

As noted above, to conduct our IRR analyses, variables that were nonnormally distributed were statistically manipulated via square root transformation, and difference scores were used instead of proportion scores to avoid missing data. Our additional analyses tested the extent to

which these statistical manipulations could have influenced our findings.

As for the square root transformation, the mean of the absolute differences between the ICCs of transformed versus non-transformed variables was .068 (*SD* = .042); the correlation between these two sets of ICCs (i.e., ICCs of transformed and ICCs of non-transformed Rorschach scores) was .922. Accordingly, one may safely conclude that using square root transformations did not notably affect our IRR estimates.

As for the effects of using difference scores, our concern was that the reliability of difference scores generally is a function of the covariance of the two composing terms. We therefore wanted to inspect whether and how the IRR of a difference score would diverge from that of its individual components. When looking at raw scores, the mean ICC was .680 (*SD* = .128) for difference scores and .801 (*SD* = .097) for their individual components. When looking at complexity-adjusted scores, the mean ICC was .664 (*SD* = .153) for difference scores and .700 (*SD* = .101) for their individual components. Thus, difference scores yielded slightly lower IRR indexes than their composing terms when using raw scores, but this phenomenon was less evident when looking at complexity-adjusted scores. A likely explanation for this finding is that because complexity is the first factor of the Rorschach, the covariance between the two individual components of a given difference score decreases when complexity-adjusted (rather than raw) scores are used. As such, the impact of the covariance between the two individual components of a given difference score on its IRR is reduced as well. In line with this hypothesis, difference scores tended to produce identical ICC values when raw or complexity-adjusted scores were used.

Finally, although our research was not designed to compare IRR of R-PAS scores for different sub-groups, our additional analyses also examined ICC values generated by our child

versus adult samples. The average ICCs were .789 (*SD* = .174) and .752 (*SD* = .156) respectively and a paired-sample *t*-test did not reveal any significant differences between the two groups ($t(59) = 1.89$, $p = .063$, $d = .25$). Thus, similar findings were observed when considering the child versus adult samples.

Discussion and Conclusions

Comparing our findings to the two previous studies, the overall ICC coefficients for raw scores were lower in the current study than those reported by Viglione et al. (2012), but similar to the findings by Kivisalu et al. (2016). The mean ICCs were .88 for Viglione et al. (2012), .78 for Kivisalu et al. (2016), and .78 for the present study, whereas the medians were .92, .80, and .80 respectively. However, Kivisalu and colleagues (2016) investigated response-level reliability of R-PAS variables, whereas Viglione and colleagues (2012) evaluated protocol-level reliability. Interestingly, variables with poor or fair reliability differ among the three studies. In our study, Vista yielded an ICC of .59: Although the ICC value was higher than those reported by Viglione et al. (2012; ICC = .44) and Kivisalu et al. (2016; ICC = .32), it was indicative of fair reliability also in our study. Vg% showed a higher ICC of .74, consistent with the findings by Kivisalu et al. (2016; Vague, ICC = .70) but different from previous R-PAS IRR findings described by Viglione et al. (2012; ICC = .54). Given that Vg% is a new code introduced in R-PAS, raters may have practiced more in coding this variable along with other new R-PAS variables, such as Synthesis (Sy; ICC = .80) or Oral Dependency Language% (ODL%; ICC = .67). Moreover, the four variables that in our study produced the lowest reliabilities (i.e., C, Mp – Ma, FQ-%, and p – a), were coded inconsistently only in the present study, while they achieved ICC ≥ .78 in Viglione et al. (2012) and ICC ≥ .73 in Kivisalu et al. (2016). Finally, while the link between base rate and ICC (i.e., the lowest ICCs were found for lowest base rate variables) was evident in

Viglione et al. (2012), in our study some of the lowest ICCs were found for variables with relatively high base rates, such as Form Quality Minus Percent (FQ-%; ICC = .53) and p – a (ICC = .54).

In our study, the variables that yielded only fair IRR may be divided into five groups: Form Dominance for color responses (i.e., C and CFC-FC), Depth determinants (i.e., FD and V), Form Quality (FQ; i.e., FQ-%, M-, WD-%, and FQu%), a newly introduced R-PAS variable [i.e., Space Reversal (SR)], and passive versus active movements (Mp – Ma and p – a). Deciding the degree of form dominance and making the distinction between FD versus Vista seem difficult to learn by students (Viglione, Meyer, Resende, & Pignolo, 2016). Moreover, Vista and C had low base rates so raters may not be experienced in coding these variables and may struggle more with learning how to code them. As for FQ, raters may have coded FQ-% and FQu% inconsistently due to having been trained in both CS and R-PAS. The R-PAS Manual provides a step-by-step method to code Form Quality (FQ), and FQ- objects are fewer in R-PAS tables than in CS tables (Meyer et al., 2011). Thus, a low ICC for FQ-% may be explained by a tendency to code FQ- derived from the CS in some of the raters, especially during the process of extrapolation to determine FQ for objects not in the FQ tables. Although most of the newly introduced R-PAS variables yielded good to excellent IRR, the ICC for SR was .59 (indicative of fair reliability). SR is one of the variables recently introduced by R-PAS and is coded when the space stands out as figure and the ink is used as background. As stated by the R-PAS Manual, identifying a threshold to evaluate whether the SR code is present is challenging, especially in complex or multidimensional responses. Thus, among the new R-PAS variables, raters may have not practice enough in differentiating SR codes from not SR codes. Finally, differentiating between active and passive movement has been recognized as a difficult challenge by different authors (e.g.,

Holaday, 1996; Viglione, 2010; Meyer et al., 2011). Although the R-PAS Manual lists a series of examples of passive/active movement thresholds, the distinction remains difficult to make, especially for human movement responses. Overall, all the variables that obtained fair IRR were those that have been described as difficult or challenging to learn and to code.

The most innovative aspect of the present study consists of having analyzed IRR for complexity-adjusted scores, which has never been done before. Complexity-adjusted scores were introduced by R-PAS to reduce extraneous variance associated with Complexity. Thus, as described in the Introduction, these scores indicate what the scores would be if the Complexity of the protocol was at a median level. Given that Complexity shares most of the common variance across test scores and that the ICC of Complexity is typically very high, we anticipated that complexity-adjusted scores would produce lower ICCs compared to raw-scores. On the one hand, our results generally confirmed this hypothesis. On the other hand, however, the differences between the ICCs of raw and complexity-adjusted scores tended to be very small, and in the majority of the cases, they did not affect the final, interpretative characterization of the ICCs.

Some limitations associated to this study deserve mentioning. First, most of the protocols were administered by graduate students, so that their ability to clarify protocol ambiguities was likely less sophisticated than it might be with expert examiners. Second, all the coders had a previous knowledge of CS, which may potentially affect our results, increasing the ICCs of R-PAS variables previously coded in CS, and decreasing ICCs of the new ones. Finally, given that three out of five coders were in training with the same mentor, our results should be replicated by inspecting IRR of raters trained by different mentors. Despite these limitations, our study has the merit to be the first to report on the IRR of R-PAS variables within an Italian context, and on the

R-PAS INTER-RATER RELIABILITY

IRR of R-PAS complexity-adjusted scores.

References

Acklin, M. W., McDowell, C. J., II, Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. *Journal of Personality Assessment, 74,* 15–47. doi:10.1207/S15327752JPA740103

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284-290. doi:10.1037/1040-3590.6.4.284

Exner, J. E. (1993). *The Rorschach: A comprehensive system. Vol. 1: Basic foundations* (3rd ed.). New York, NY: Wiley.

Exner, J. E. (2003). *The Rorschach: A comprehensive system. Vol. 1: Basic foundations and principles of interpretation* (4th ed.). Hoboken, NJ: Wiley.

Giromini, L., Viglione, D. J., & McCullaugh, J. (2015). Introducing a Bayesian approach to determining degree of fit with existing Rorschach norms. *Journal of Personality Assessment, 97*, 354-363. doi:10.1080/00223891.2014.959127

Holaday, M. (1996). Coding and interpreting movement on the Rorschach. *Assessment, 3*(2), 103-110. doi: 10.1177/107319119600300201

Kivisalu, T. M., Lewey, J. H., Shaffer, T. W., & Canfield M. L. (2016). An investigation of interrater reliability for the Rorschach Performance Assessment System (R-PAS) in a nonpatient U.S. sample. *Journal of Personality Assessment, 98*(4), 382-390. doi:10.1080/00223891.2015.1118380

McDowell, C. J., & Acklin, M. W. (1996). Standardizing procedures for calculating Rorschach interrater reliability: Conceptual and empirical foundations. *Journal of Personality Assessment, 66,* 308–320. doi:10.1207/s15327752jpa6602_9

Meyer, G. J. (2004). The reliability and validity of the Rorschach and TAT compared to other psychological and medical procedures: An analysis of systematically gathered evidence. In M. Hersen (Ed.-in-Chief) & M. Hilsenroth & D. Segal (Eds.), *Comprehensive handbook of psychological assessment: Vol. 2. Personality assessment* (pp. 315–342). Hoboken, NJ: Wiley.

Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment, 13,* 486–502. doi:10.1037/1040-3590.13.4.486

Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *A manual for the Rorschach Performance Assessment System*. Toledo, OH: R–PAS.

Meyer, J. G., Viglione, D. J., & Giromini, L. (2014). An Introduction to Rorschach-Based Performance Assessment. In R. P. Archer, & S. R. Smith, (Eds.) *Personality Assessment* (pp. 301-370). Mahwah, NJ: Routledge.

Mihura, J. L., Roy, M., & Graceffo, R. A. (2016). Psychological assessment training in clinical psychology doctoral programs. *Journal of Personality Assessment*. Advance online publication. doi:10.1080/00223891.2016.1201978.

Reese, J. B., Viglione, D. J., & Giromini, L. (2014). A comparison between Comprehensive System and an early version of the Rorschach Performance Assessment System administration with outpatient children and adolescents. *Journal of Personality Assessment, 96,* 515-522. doi:10.1080/00223891.2014.889700

Shrout, P. E., & Fliess, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428. doi:10.1037/0033-2909.86.2.420

Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment, 11,* 251–265. doi:10.1037/1040-3590.11.3.251

Viglione, D. J. (2010). Rorschach coding solutions: A reference guide for the Comprehensive System (2nd ed.). San Diego, CA: Author.

Viglione, D. J., & Giromini, L. (2016). The effects of using the International versus Comprehensive System norms for children, adolescents, and adults. *Journal of Personality Assessment, 98,* 391-397. doi:10.1080/00223891.2015.1136313

Viglione, D. J., & Meyer, G. J. (2008). An overview of Rorschach psychometrics for forensic research. In C. B. Gacono, F. B. Evans, N. Kaser-Boyd, & L. A. Gacono (Eds.), *The handbook of forensic Rorschach assessment* (21–53). New York, NY: Routledge/Taylor & Francis.

Viglione, D. J., & Taylor, N. (2003). Empirical support for intercoder reliability of Rorschach Comprehensive System coding. *Journal of Clinical Psychology, 59,* 111–121. doi:10.1002/jclp.10121

Viglione, D. J., Blume-Marcovici, A. C., Miller, H. L., Giromini, L., & Meyer, G. (2012). An Inter-Rater Reliability Study for the Rorschach Performance Assessment System. *Journal of Personality Assessment, 94*(6), 607-612. doi:10.1080/00223891.2012.684118

Viglione, D. J., Meyer, G. J., Jordan, R. J., Converse, G. L., Evans, J., MacDermott, D., & Moore, R. C. (2015). Developing an alternative Rorschach administration method to optimize the number of responses and enhance clinical inferences. *Clinical Psychology and Psychotherapy, 22,* 546-558. doi:10.1002/cpp.1913

Viglione, D. J., Meyer, G. J., Resende, A. C., & Pignolo, C. (2016). A survey of challenges experienced by new learners coding the Rorschach. *Journal of Personality Assessment*.

Advance online publication. doi:10.1080/00223891.2016.1233559

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56 –75). Thousand Oaks, CA: Sage.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7,* 3–10. doi:10.1111/j.1467-9280.1996.tb00658.x

Table 1. Inter-rater reliabilities for R-PAS Summary Scores on Page 1 and Page 2

	Raw scores		Complexity-Adjusted Scores		
Variable	ICC	Classification	ICC	Classification	Base Rate
Page 1					
Administration Behaviors & Observations					
Pr	.81	Excellent			Rare
Pu	.78	Excellent			Rare
CT	.98	Excellent			Common
Engagement & Cognitive Processing					
Complexity	.94	Excellent			Common
R	1.00	Excellent	.96	Excellent	Common
F%	.93	Excellent	.87	Excellent	Common
Blend	.86	Excellent	.71	Good	Common
Sy	.80	Excellent	.61	Good	Common
MC	.90	Excellent	.63	Good	Common
MC - PPD	.80	Excellent	.79	Excellent	Common
M	.92	Excellent	.79	Excellent	Common
M - WSumC	.86	Excellent	.85	Excellent	Common
CFC - FC	.57	Fair	.58	Fair	Infrequent
Perception & Thinking Problems					
EII-3	.76	Excellent	.74	Good	Common
TP-Comp	.64	Good	.63	Good	Common
WSumCog	.77	Excellent	.81	Excellent	Common
SevCog	.65	Good	.67	Good	Rare
FQ-%	.53	Fair	.51	Fair	Common
WD-%	.58	Fair	.59	Fair	Common
FQo%	.82	Excellent	.79	Excellent	Common
P	.84	Excellent	.83	Excellent	Common
Stress & Distress					
m	.65	Good	.64	Good	Infrequent
Y	.77	Excellent	.69	Good	Infrequent
MOR	.78	Excellent	.75	Excellent	Infrequent
SC-Comp	.75	Excellent	.71	Good	Common
Self & Other Representation					
ODL%	.67	Good	.63	Good	Common
SR	.59	Fair	.61	Good	Infrequent
MAP - MAH	.71	Good	.71	Good	Rare
PHR - GHR	.72	Good	.71	Good	Common
M-	.57	Fair	.57	Fair	Rare
AGC	.64	Good	.62	Good	Common
V-Comp	.93	Excellent	.90	Excellent	Common
H	.83	Excellent	.81	Excellent	Common

	Raw scores		Complexity-Adjusted Scores		
Variable	ICC	Classification	ICC	Classification	Base Rate
COP	.84	Excellent	.80	Excellent	Infrequent
MAH	.80	Excellent	.74	Excellent	Rare
Page 2					
Engagement & Cognitive Processing					
W%	.97	Excellent	.95	Excellent	Common
Dd%	.83	Excellent	.81	Excellent	Common
SI	.77	Excellent	.69	Good	Common
IntCont	.79	Excellent	.76	Excellent	Common
Vg%	.74	Excellent	.77	Excellent	Common
V	.59	Fair	.59	Fair	Rare
FD	.57	Fair	.55	Fair	Infrequent
R8910%	.95	Excellent	.95	Excellent	Common
WSumC	.83	Excellent	.75	Excellent	Common
C	.41	Fair	.41	Fair	Rare
Mp - Ma	.51	Fair	.43	Fair	Infrequent
Perception & Thinking Problems					
FQu%	.59	Fair	.55	Fair	Common
Stress & Distress					
PPD	.89	Excellent	.83	Excellent	Common
YTVC'	.91	Excellent	.86	Excellent	Common
CBlend	.79	Excellent	.71	Good	Rare
C’	.88	Excellent	.85	Excellent	Common
CritCont%	.93	Excellent	.92	Excellent	Common
Self & Other Representation					
SumH	.95	Excellent	.88	Excellent	Common
NPH - H	.74	Excellent	.78	Excellent	Common
r	.98	Excellent	.97	Excellent	Rare
p - a	.54	Fair	.47	Fair	Common
AGM	.82	Excellent	.85	Excellent	Rare
T	.87	Excellent	.87	Excellent	Rare
PER	.82	Excellent	.83	Excellent	Rare
An	.96	Excellent	.95	Excellent	Infrequent

Notes. Bolded = square-root transformed variable; Underlined = proportion variable computed as subtraction.

Table 2. Summary of intraclass correlation inter-rater reliability results for 60 Rorschach Performance Assessment System variables

	Raw scores	Complexity Adjusted Scores
M	.78	.74
SD	.14	.14
Minimum	.41	.41
25th percentile	.66	.63
Median	.80	.75
75th percentile	.88	.84
Maximum	1.00	.97
No. of poor ICCs < .40	0	0
No. of fair ICCs .40-.59	11 (18%)	10 (18%)
No. of good ICCs .60-.74	7 (12%)	16 (29%)
No. of excellent ICCs >= .75	42 (70%)	30 (54%)
Mean ICC for 13 low base rate, rare variables	.74	.72
Mean ICC for 9 moderately low base rate, infrequent variables	.69	.67
Mean ICC for 38 common base rate variables	.81	.76

R-PAS INTER-RATER RELIABILITY

Table 3. Descriptive statistics and base rates

	Rater 1				Rater 2				
Variable	N	Range	M	SD	N	Range	M	SD	BR
Page 1									
Administration Behaviors & Observations									
Pr	112	0 – 7	.71	1.10	112	0 – 7	.71	1.13	.71
Pu	112	0 – 5	.33	.82	112	0 – 5	.30	.79	.32
CT	112	0 – 21	5.07	5.19	112	0 – 22	4.90	5.26	4.99
Engagement & Cognitive Processing									
Complexity	112	33 – 164	80.52	24.87	112	37 – 141	79.29	24.53	79.91
R	112	18 – 36	25.46	4.16	112	18 – 36	25.47	4.19	25.47
F%	112	4 – 89	41.07	19.01	112	0 – 89	40.10	18.93	40.58
Blend	112	0 – 15	5.02	3.83	112	0 – 20	4.89	4.08	4.96
Sy	112	0 – 17	6.77	4.00	112	0 – 17	6.13	3.88	6.45
MC	112	0 – 16.5	7.20	3.94	112	0 – 16	7.02	3.73	7.11
MC - PPD	112	4 – 39	18.88	9.00	112	4 – 51	18.95	8.92	18.92
M	112	0 – 12	3.88	2.75	112	0 – 11	3.87	2.69	3.87
M-WSumC	112	-8 – 9	.55	3.17	112	-7 – 8	.71	2.94	3.87
CFC-FC	112	-11 – 9	-.31	2.77	112	-8 – 7	-.50	2.60	1.88
Perception & Thinking Problems									
EII-3	112	28 – 172	71.54	30.15	112	29 – 199	69.59	28.34	70.56
TP-Comp	112	34 – 130	63.15	16.98	112	36 – 110	63.30	15.74	63.23
WSumCog	112	0 – 76	11.21	13.07	112	0 – 51	11.34	12.14	11.27
SevCog	112	0 – 11	.98	1.83	112	0 – 7	.88	1.53	.93
FQ-%	112	0 – 47	15.77	10.97	112	0 – 65	15.97	11.58	15.87
WD-%	112	0 – 44	13.91	9.93	112	0 – 56	14.43	10.81	14.17
FQo%	112	21 – 82	48.72	13.70	112	20 – 82	50.08	14.71	49.40
P	112	1 – 11	5.09	2.12	112	1 – 12	5.19	2.15	5.14
Stress & Distress									
m	112	0 – 8	1.79	1.61	112	0 – 8	1.88	1.65	1.83
Y	112	0 – 8	1.90	1.95	112	0 – 8	1.97	1.97	1.94
MOR	112	0 – 11	2.13	2.05	112	0 – 8	1.62	1.64	1.88
SC-Comp	112	98 – 220	151.33	28.42	112	89 – 202	150.27	28.96	150.80
Self & Other Representation									
ODL%	112	0 – 37	8.10	7.91	112	0 – 29	9.04	7.55	8.57
SR	112	0 – 11	1.09	1.54	112	0 – 12	1.19	1.60	1.14
MAP - MAH	112	-5 – 3	-.40	1.38	112	-4 – 6	-.38	1.54	.57
PHR - GHR	112	-11 – 8	-.54	3.91	112	-8 – 7	-.56	3.36	3.45
M-	112	0 – 5	.64	1.04	112	0 – 4	.64	.96	.64
AGC	112	0 – 10	2.59	2.37	112	0 – 10	3.13	2.28	2.86
V-Comp	112	85 – 183	135.54	22.69	112	.8 – 7.5	131.46	20.92	133.50

R-PAS INTER-RATER RELIABILITY

	Rater 1				Rater 2				
Variable	N	Range	M	SD	N	Range	M	SD	BR
H	112	0 – 9	2.61	2.16	112	0 – 8	2.45	2.00	2.53
COP	112	0 – 6	1.27	1.40	112	0 – 5	1.11	1.24	1.19
MAH	112	0 – 5	.93	1.16	112	0 – 5	.99	1.17	.96
Page 2									
Engagement & Cognitive Processing									
W%	112	7 – 89	41.30	20.74	112	7 – 88	40.51	19.66	40.91
Dd%	112	0 – 50	15.03	8.98	112	0 – 45	14.96	8.87	14.99
SI	112	0 – 11	3.19	2.25	112	0 – 13	2.60	2.40	2.89
IntCont	112	0 – 9	2.45	2.56	112	0 – 9	2.23	2.46	2.34
Vg%	112	0 – 26	4.18	5.94	112	0 – 33	3.70	6.60	3.94
V	112	0 – 7	.93	1.34	112	0 – 8	.91	1.53	.92
FD	112	0 – 6	1.21	1.31	112	0 – 8	1.12	1.64	1.17
R8910%	112	17 – 42	30.78	4.11	112	21 – 42	30.91	3.93	30.84
WSumC	112	0 – 12	3.32	2.29	112	0 – 8	3.15	2.02	3.24
C	112	0 – 6	.49	1.08	112	0 – 4	.34	.73	.42
Mp - Ma	112	-6 – 7	-.50	2.14	112	-6 – 7	-.03	2.14	1.82
Perception & Thinking Problems									
FQu%	112	8 – 65	34.13	12.04	112	0 – 65	32.03	12.28	33.08
Stress & Distress									
PPD	112	1 – 27	11.69	6.35	112	2 – 40	11.93	6.31	11.81
YTVC'	112	0 – 23	6.70	4.63	112	0 – 25	6.61	4.70	6.65
CBlend	112	0 – 6	1.08	1.31	112	0 – 6	.91	1.31	1.00
C’	112	0 – 12	3.37	2.58	112	0 – 16	3.16	2.68	3.26
CritCont%	112	0 – 111	22.74	18.51	112	0 – 117	20.73	17.28	21.74
Self & Other Representation									
SumH	112	0 – 16	6.44	3.03	112	0 – 15	6.63	3.17	6.53
NPH - H	112	-7 – 10	1.22	3.23	112	-6 – 10	1.73	3.27	4.00
r	112	0 – 4	.47	.94	112	0 – 5	.49	1.05	.48
p - a	112	-10 – 11	-.24	3.50	112	-9 – 10	.26	3.64	4.55
AGM	112	0 – 5	.91	1.20	112	0 – 5	.80	1.07	.86
T	112	0 – 6	.50	.97	112	0 – 7	.56	1.16	.53
PER	112	0 – 10	.92	1.61	112	0 – 8	.91	1.50	.92
An	112	0 – 7	1.69	1.70	112	0 – 7	1.68	1.70	1.68

Note. Underlined = proportion variable computed as subtraction.